
Désaisonnalisation de séries longues : l'apport des modèles structurels

Anna SMYK ()*

() Insee, Département des Méthodes Statistiques*

`anna.smyk@insee.fr`

Mots-clés. : X-13 Arima, Tramo-Seats, Basic Structural Models, Seasonality break, JDemetra+

Domaines. Séries temporelles, Désaisonnalisation

Résumé

Le processus générateur des données évolue au cours du temps, en même temps que les institutions et les comportements. Ainsi, désaisonnaliser des séries de plus de vingt ans est considéré comme sous-optimal par les Guidelines d'Eurostat sur l'ajustement saisonnier. Les algorithmes de désaisonnalisation les plus courants, X-13 Arima et Tramo-Seats, reposent, au moins en partie, sur l'estimation d'effets calculés sur l'ensemble de la période, ce qui peut constituer une faiblesse pour le traitement des séries longues.

Face à ces limitations, nous présentons une autre classe de méthodes : les modèles structurels. Ceux-ci permettent une modélisation explicite et flexible des composantes inobservables d'une série temporelle et semblent plus adaptés au traitement d'un processus à structure évolutive. Nous mettons notamment en oeuvre une méthode de traitement des ruptures structurelles dans le profil saisonnier, utilisant les modèles structurels multi-variés.

La désaisonnalisation de séries longues expose l'utilisateur à une estimation erronée du message conjoncturel en fin de période, élément souvent crucial pour les utilisateurs. A travers quelques exemples caractéristiques nous étudions la stabilité de celui-ci selon l'algorithme utilisé et concluons qu'une approche non paramétrique de la décomposition (X-13) reste de ce point de vue plus fiable qu'une approche paramétrique, y compris avec les modèles structurels.

Pour l'ensemble des calculs, nous utilisons les algorithmes disponibles dans le logiciel JDemetra+, qui, en plus de proposer les algorithmes X-13 et Tramo-Seats, permet un ajustement saisonnier avec les modèles structurels, comprenant les représentations les plus utilisées de la composante saisonnière ainsi qu'une correction des effets de calendrier et des points atypiques.

Abstract in English

Performing seasonal adjustment on long time series, longer than 20 years, is considered sub-optimal et stated as such by the Eurostat guidelines on seasonal adjustment. Indeed most popular algorithms (X-13-Arima et Tramo-Seats) are partly or completely based on a stable modelisation assumption of the whole series span, which is not realistic as data generating process evolves with behavioral and institutional changes. For this reason, we investigate another class of methods based on the state space framework (SSF) and its recursive estimation method : the basic structural models (BSM), which are especially proficient when dealing with time series displaying structural breaks. We provide some empirical insights on the stability across algorithms of a series message delivered to policy makers, using JDemetra+ 3.0 software.

Introduction

Les producteurs de données cvs-cjo sont le plus souvent confrontés aux problèmes inhérents à la désaisonnalisation de séries trop courtes (instabilité des estimations, voire impossibilité d'utiliser les algorithmes usuels), apparaissant avec chaque nouvel indicateur mis en place pour répondre à un nouveau besoin de mesure. Mais les séries (très) longues font également partie du paysage dans les institutions statistiques et peuvent de surcroît présenter des ruptures, dues notamment à des changements de concept ou méthode.

Il est généralement admis que le processus générateur des données évolue au cours du temps, en même temps que les institutions et les comportements. Les algorithmes de désaisonnalisation courants reposent pour partie sur l'estimation d'effets calculés sur l'ensemble de la période, avec toutefois d'importantes différences d'approche, que nous allons examiner.

Pour cette raison, désaisonnaliser des séries longues, de plus de 20 ans, peut des donner des estimations sous optimales en début et surtout en fin de période, souvent le point plus importante pour l'utilisateur. Des études[1] montrent que le message conjoncturel délivré en fin de période peut-être sensiblement modifié lorsque l'on ramène une période d'estimation de plus de vingt ans à une dizaine d'années et la qualité de la désaisonnalisation, notamment du pré-ajustement s'en trouve améliorée. Aussi, les guidelines d'Eurostat sur l'ajustement saisonnier [2] recommandent une désaisonnaliser par sous-périodes.

Si toutefois, on ne souhaite pas raccourcir ses séries, il est naturel de se tourner vers une classe de méthodes plus flexibles, ne reposant pas sur des hypothèses de comportements constants : les modèles structurels. Le recours à ceux-ci est moins fréquent dans la pratique de l'ajustement saisonnier des statistiques officielles. Pourtant, leur estimation récursive reposant sur le filtre de Kalman en fait une alternative légitime lorsqu'il s'agit de traiter le changement structurel dans le temps. En nous plaçant dans le cadre des "Basic structural models" (BSM) tel que défini par Harvey (1989)[3], nous nous proposons d'examiner dans quelle mesure les modèles structurels offrent des solutions aux problèmes spécifiques posés par les séries longues : saisonnalité et effets de calendrier évolutifs, ruptures du profil saisonnier.

La première partie de l'article présente les enjeux spécifiques à la désaisonnalisation des séries longues à travers les caractéristiques des deux algorithmes les plus utilisés pour l'ajustement saisonnier des statistiques officielles, X-13 et Tramo-Seats. Dans la seconde partie, on se tourne vers un cadre alternatif, les modèles structurels, avec une attention particulière pour le traitement des ruptures de série. La troisième partie présente quelques faits stylisés quant à l'impact du choix de l'algorithme sur la stabilité du message conjoncturel en fin de période.

1 Les enjeux spécifiques à la désaisonnalisation de séries longues

1.1 Décomposer pour désaisonnaliser

L'ajustement saisonnier est utilisé pour purger les séries de mouvements périodiques, afin d'éviter que les variations conjoncturelles et les tendances de fond ne soient masquées par ces phénomènes récurrents, souvent de grande ampleur. Pour cette raison, les indicateurs économiques sont publiés corrigés des variations saisonnières (cvs) et des effets de calendrier (cjo).¹

Désaisonnaliser nécessite d'estimer une composante saisonnière capturant l'ensemble des mouvements périodiques et contenant donc un éventuel effet de calendrier, afin de l'enlever de la série brute $Y_{cvs} = Y - S$. Pour ce faire, on décompose la série brute en trois composantes inobservables : saisonnalité, tendance et irrégulier ($Y = S + T + I$), où la tendance T représente les mouvements de long terme et I le bruit engendré par des événements exceptionnels (grèves, conditions climatiques, chocs de court terme etc.) et l'ensemble des erreurs de mesure. Il existe une infinité de décompositions possibles : le choix de la solution dépend de la définition des composantes, propre à chaque algorithme et à l'origine de leurs différences.

1.2 Les algorithmes les plus couramment utilisés

Les deux méthodes les plus utilisées actuellement au sein des instituts nationaux de statistique européens et des banques centrales sont Tramo-Seats (paramétrique) et X13-Arima (semi-paramétrique), ce dernier étant quasi exclusivement utilisé au sein du service statistique public français. Ils sont disponibles dans le logiciel JDemetra+ [4] recommandé par Eurostat et la banque centrale européenne, *via* une interface graphique ou le package RJDemetra[5].

1.3 Limitations pour désaisonnaliser les séries longues

Ces méthodes procèdent en deux étapes pour désaisonnaliser : pré-ajustement de la série brute puis décomposition.

1.3.1 Pré-ajustement

Afin d'améliorer la décomposition, la série brute est linéarisée en enlevant les effets de calendrier ou les points atypiques. Les effets de calendrier seront définitivement retirés de la série *cvs* alors que les effets des points atypiques affectant l'irrégulier ou la tendance seront réinjectés dans la série finale.² Dans la pratique, les phases de pré-ajustement sont très proches dans les deux algorithmes et reposent sur une modélisation Reg-Arima. Dans ce cadre, les effets estimés et la structure d'autocorrélation des résidus sont supposés constants sur l'ensemble de la période d'estimation. Cette hypothèse peut se révéler très forte dans le cas de séries de plus de 20 ans, car le processus générateur de données a des chances importantes d'avoir substantiellement évolué dans le temps.

1. Dans la suite du texte, nous ferons le raccourci usuel en désignant par "corrigées des variations saisonnières" (cvs) des séries aussi, éventuellement corrigées des jours ouvrables (cvs-cjo)

2. S'ils affectent S , cas plus rare, ils seront aussi enlevés de la série *cvs*

1.3.2 Décomposition

Les philosophies de décomposition sont quant à elles très différentes d'un algorithme à l'autre. X-13 repose sur une utilisation itérative de moyennes mobiles, avec l'une d'entre elles dont l'ordre est égal à la périodicité de la série et qui permet d'éliminer la saisonnalité [6]. Par rapport à la méthode historique X-11, développée dans les années 50, la structure globale du filtrage a peu évolué mais le processus a été enrichi en utilisant l'apport des modèles Arima. Les propriétés des filtres symétriques étant meilleures pour estimer la série *cvs*, une prévision de la série brute a été introduite, celle-ci reposant sur le modèle Arima ajusté dans l'étape de linéarisation. Ce changement fait désormais dépendre d'un modèle estimé sur l'ensemble de la période les valeurs de la fin de la série, alors que le filtrage lui évolue par fenêtre mobile. Cette dépendance, d'ailleurs très débattue [7], apparaît comme un point faible pour désaisonnaliser les séries longues alors même que le filtrage par fenêtre mobile semble un atout pour s'adapter à un processus évolutif.

Pour décomposer la série brute SEATS[8] se fonde quant à lui sur le modèle Arima ajusté par Tramo sur l'intégralité de la série brute, pour modéliser, là aussi sous forme Arima, les composantes T , S et I et déterminer les filtres qui permettront de les extraire[9]. Ceux-ci sont donc fixés au vu de l'ensemble de la période, contrairement au filtrage de X-13.

Le caractère évolutif du processus générateur de données ne pouvant être complètement pris en compte par les algorithmes usuels, désaisonnaliser les séries de plus de 20 ans avec un paramétrage unique est classé comme alternative C, à éviter, dans les guidelines d'Eurostat sur l'ajustement saisonnier. Il y est recommandé de les ajuster par sous-périodes, longues d'au moins 7 ans. Cela nécessite de déterminer des sous-périodes adéquates et une méthode de raccordement, ce qui occasionne un coût supplémentaire pour le producteur[1]

Toutefois, au delà de ces deux algorithmes largement utilisés, il existe une alternative qualifiée de "raisonnable" par les Guidelines d'Eurostat et fondée sur les modèles espaces d'états³ dont on peut se demander s'ils permettant de mieux traiter les problèmes spécifiques posés par les séries longues, sans les couper.

2 Un cadre alternatif pour la désaisonnalisation

2.1 Les modèles structurels

Le cadre général de modèles "espaces d'états" permet d'estimer des composantes inobservables d'une variable mesurée (équation de mesure) en modélisant explicitement la dynamique sous-jacente de celles-ci (équations d'état ou de transition) et en utilisant le filtre de Kalman comme détaillé par Gouriéroux et Montfort(1995) [10]. Ce cadre statistique se prête bien à la désaisonnalisation qui est également fondée sur l'estimation de composantes inobservables. Il a notamment été formalisé par Harvey (1989)[3] qui a introduit les modèles structurels en séries temporelles (basic structural models ou BSM). La série brute y_t , contrairement à la modélisation Arima, y est explicitement modélisée en fonction de ses composantes :

$$y_t = \mu_t + \gamma_t + X_t\beta + \epsilon_t, \quad t = 1, \dots, n,$$

où

3. "unobserved component methods (...) based on state space models represent a reasonable alternative, provided they allow for a complete calendar et outlier treatment"

μ_t = tendance linéaire locale
 γ_t = composante saisonnière de période s
 X_t = variables explicatives (calendrier, outliers)
 ϵ_t = irrégulier

— Matrice d'état :

$$\alpha_t = (\epsilon_t, \mu_t, \beta_t, \gamma_{i_t, t}, \dots, \gamma_{j_t, t})'$$

— Équation de mesure :

$$y_t = Z\alpha_t \quad \text{où} \quad Z = (1, 1, 0, 1, 0, \dots, 0)$$

— Équation de transition (définie par blocs) :

$$\begin{aligned} \epsilon_{t+1} &= 0\epsilon_t + \zeta_{\epsilon t} \sim N(0, \sigma_{\epsilon}^2) \\ \begin{bmatrix} \mu_{t+1} \\ \beta_{t+1} \end{bmatrix} &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + \begin{bmatrix} \zeta_{\mu t} \\ \zeta_{\beta t} \end{bmatrix} \quad \text{où} \quad \begin{cases} \zeta_{\mu t} \sim N(0, \sigma_{\mu}^2) \\ \zeta_{\beta t} \sim N(0, \sigma_{\beta}^2) \end{cases} \\ \begin{bmatrix} \gamma_{i_{t+1}t+1} \\ \vdots \\ \vdots \\ \gamma_{j_{t+1}t+1} \end{bmatrix} &= \begin{bmatrix} 0 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 \\ -1 & -1 & \dots & -1 \end{bmatrix} \begin{bmatrix} \gamma_{i_t t} \\ \vdots \\ \vdots \\ \gamma_{j_t t} \end{bmatrix} + \zeta_{\gamma t} \end{aligned}$$

$$\begin{aligned} o: \quad \zeta_{\gamma t} &\sim N(0_{s-1}, \sigma_{\gamma}^2 \Omega) \\ i_t &\text{ is the (0-based) period of } t, \quad j_t = i_t + s - 2 \pmod s \end{aligned}$$

La composante saisonnière peut-être modélisée de différentes manières [11]. Elle est représentée ici comme une marche aléatoire multivariée et différentes matrices de variance-covariance de l'erreur Ω peuvent être spécifiées pour obtenir les représentations usuelles (trigonométrique, Harrison-Stevens, dummy). L'ensemble des effets déterministes (outliers et calendrier) est représenté directement dans l'équation de mesure par le terme $X\beta$, la linéarisation et la décomposition étant ici simultanées contrairement à Tramo-Seats ou X13. La détection d'outliers se fait sur les résidus lissés des différentes composantes du modèle[12]. Cette modélisation est fréquemment utilisée dans la littérature pour estimer des effets de calendrier variables, comme le détaille l'article du Handbook on seasonal adjustment sur ce thème[13]

Les BSM ne sont pas, ou peu, utilisés pour la production de statistiques officielles. Pourtant Durbin[14] les qualifie de plus transparents, flexibles et stables que la modélisation Arima et pense que leur non utilisation n'est due qu'à leur méconnaissance dans les instituts statistiques. Ils ont en effet l'avantage de permettre une modélisation explicite des composantes, là où un modèle Arima apparaît plus comme une boîte noire. Dans ce dernier, la tendance et la saisonnalité sont éliminées par différenciation car il est nécessaire de stationnariser les séries pour mettre oeuvre la modélisation et une méthode supplémentaire est nécessaire pour obtenir le résultat recherché, c'est à dire une décomposition. Il montre que la modélisation structurelle est équivalente à une modélisation Arima airline

(0,1,1)(0,1,1). La modélisation Arima classique, utilisée dans les algorithmes X-13 Arima (pré-ajustement) et Tramo-Seats (pré-ajustement et décomposition), est homogène dans le temps ce qui apparaît comme une faiblesse pour le traitement des séries longues. Les BSM offrent une flexibilité due au caractère récursif de l'estimation et se prêtent donc aux changements de structure et coefficients variables.

2.2 Les séries avec rupture de profil saisonnier

Le traitement des outliers tel qu'il est possible de le mettre en oeuvre avec les modèles structurels offre la possibilité de détecter et de modéliser une rupture soudaine dans l'ensemble du profil saisonnier de la série (seasonality break). Le seasonal outlier (SO) disponible dans les algorithmes classiques ne permet de modéliser un choc que sur l'une des périodes, celui-ci étant compensé par un ajustement équivalent de tous les autres types de période. Pour les séries longues, la modélisation d'un choc de type global semble pertinente dès que la série a subi un changement de concept dans la mesure et devient une superposition de deux sous séries. Dans ce cas, il est raisonnable de désaisonnaliser ces sous-périodes séparément et de recoller les résultats. La modélisation BSM proposée ici permet d'éviter de tels problèmes de raccordement. Dans un premier temps, les résidus lissés (après application du "kalman filter et smoother", KFS) de l'ensemble des composantes saisonnières permettent de détecter la rupture selon la méthode présentée par Palate *et al.* [12]. Dans un second temps, on considère que la série avant et après la rupture correspond à deux séries qui ont une tendance et un irrégulier communs mais des saisonnalités différentes. La modélisation multivariée, ci-dessous permet d'obtenir *in fine* une série cvs sur l'ensemble de la période :

- Forme uni-variée (équation de mesure)

$$y_t = T_t + 1_{t < t_b} S_{1,t} + 1_{t \geq t_b} S_{2,t} + \nu_t$$

- Forme multi-variée équivalente

$$y_{1,t} = T_t + S_{1,t} + \nu_t$$

$$y_{2,t} = T_t + S_{2,t} + \nu_t$$

avec

$$y_{1,t} = \begin{cases} y_t & \text{if } t < t_b \\ \text{na} & \text{if } t \geq t_b \end{cases} \quad y_{2,t} = \begin{cases} \text{na} & \text{if } t < t_b \\ y_t & \text{if } t \geq t_b \end{cases}$$

La tendance et la série cvs sont ici estimées sur l'ensemble de la période, comme l'illustre l'exemple ci-dessous.

2.3 Exemple de rupture du profil saisonnier

La production de sucre en Belgique connaît une rupture majeure du profil saisonnier à la fin de l'année 1986, suite à une réorganisation de la production (figure 1).

Si l'on essaie d'obtenir une cvs avec Tramo-Seats par exemple, et ce, même en auto-risant les Seasonal Outliers (SO), on génère un grand nombre d'outliers et une rupture

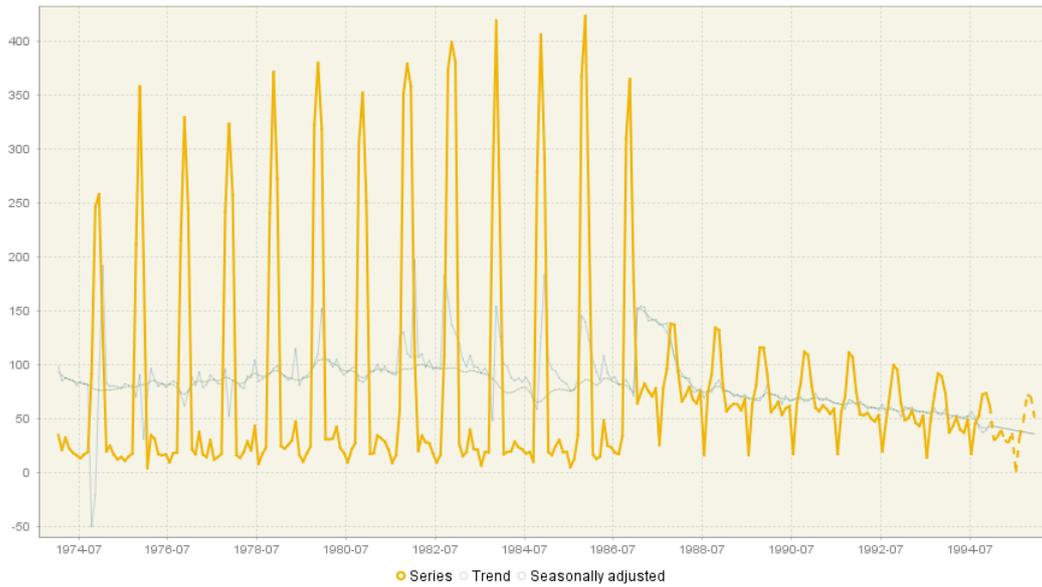


FIGURE 1 – Production de sucre en Belgique série brute

dans la tendance, comme le montre la figure 2. Si l'on opte pour un ajustement saisonnier utilisant l'approche BSM multi-variée, on parvient à une tendance et série *cvs* plus satisfaisante sur l'ensemble de la période.

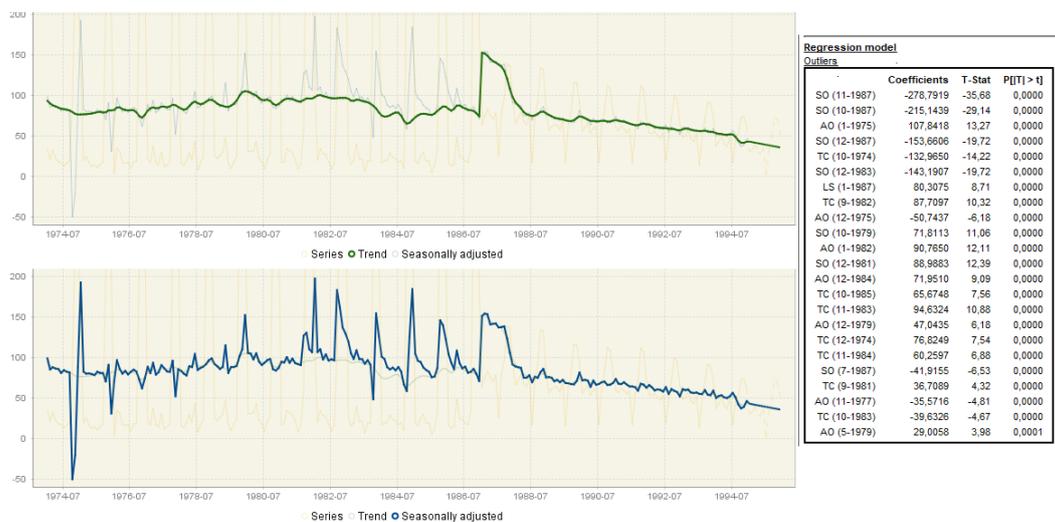


FIGURE 2 – Tendence (haut) et série cvs (bas) obtenues avec Tramo-Seats

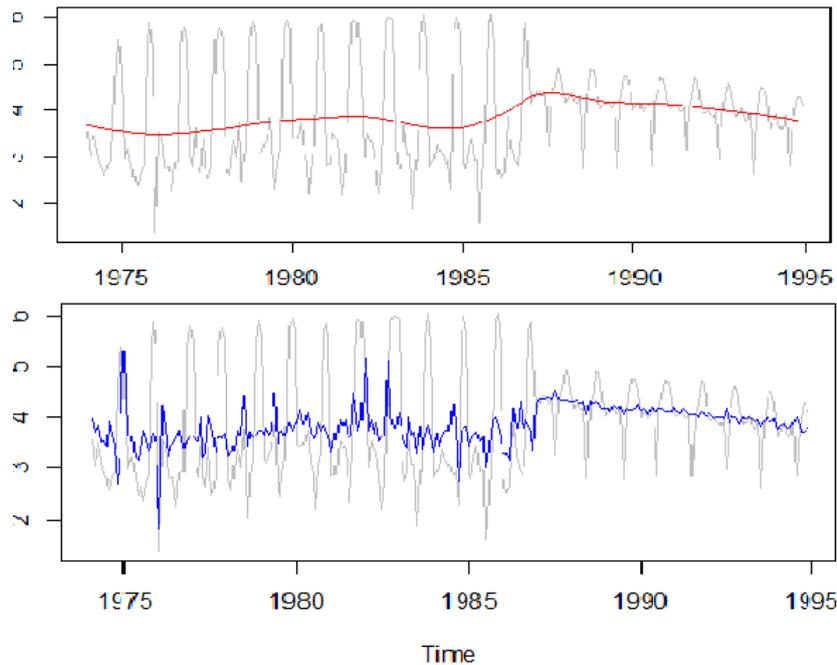


FIGURE 3 – Tendance (en haut) et série cvs (en bas) obtenues avec un BSM multi-varié

3 Illustration empirique : stabilité du message en fin de série

On se demande désormais quelle est la méthode la plus coûteuse quand on laisse ses séries vieillir sans les désaisonnaliser par sous-période, processus qui a d'ailleurs un certain coût ne serait ce que pour le choix d'une stratégie de raccordement [1]. Les études sur l'influence de la longueur de la période d'estimation sur le message conjoncturel délivré en fin de période mettent en évidence de fortes différences, qui plaident pour une coupure de séries. Toutefois *Mazzi et al.* (2003)[15] concluent que, de ce point de vue, avoir des séries trop courtes est encore plus dommageable. De manière attendue, la méthode X-13 apparaît plus adaptée aux séries longues que Tramo-Seats pour *Webel*(2016)[16], ce qui est confirmé par *Buono et al.*(2018) [17].

Si l'on accepte le consensus de la littérature selon lequel la longueur optimale d'estimation est de 8 à 12 ans[17], on peut regarder dans quels cas, selon le choix l'algorithme et le profil de la série, la pénalité de faire tout de même l'estimation sur des séries de plus de vingt ans est la plus élevée. En introduisant des distances appropriées, on compare, pour chaque méthode, l'ajustement d'une série de longueur optimale à sa version longue. Nous mesurons ici les différences en niveau, en taux de croissance et les changements de sens de variation sur la période optimale. Ces indicateurs sont également recalculés sur les trois dernières années, qui est la période d'intérêt pour les conjoncturistes.

Dans les données utilisées (Enquête de conjoncture dans l'industrie et Indice à la production industrielle publiés par l'Insee), la version longue, notée cvs^L correspond à la période 1990-2019 et la version optimale, notée cvs^O à 2012-2019. Nous coupons fin 2019 afin de nous affranchir d'une période unique dans l'histoire des séries dont l'étude dépasse

le cadre cet article.

3.1 Indicateurs de comparaison

On mesure la différence entre la série estimée sur longue période (1990-2019) et sur période "optimale" (2012-2019) avec trois indicateurs : la moyenne des écarts en niveau (sur la période optimale et sur les trois dernières années), la moyenne des écarts en taux d'évolution et la proportion de points pour lesquels les sens de variation des deux *cvs* sont différents, indépendamment du taux de variation.

3.1.1 Écarts en niveau

Pour une série brute Y_t , $t \in \{1/2012 \dots 12/2019\}$ ou $t \in \{1/2017 \dots 12/2019\}$, la moyenne des écarts en niveau est calculée ainsi :

$$EN(Y_t) = \frac{1}{T} \sum_{k=1}^T \left| \frac{cvs_k^L - cvs_k^O}{cvs_k^L} \right| \times 100$$

Un écart important en niveau peut être engendré par une différence ponctuelle, alors que les dynamiques sont par ailleurs proches, on compare donc aussi les taux d'évolution.

3.1.2 Différences sur les taux d'évolution

La moyenne des différences d'évolution entre les deux versions de la série désaisonnalisée est calculée ainsi :

$$ETC(Y_t) = \frac{1}{T-1} \sum_{k=2}^T \left| \frac{cvs_k^L - cvs_{k-1}^L}{cvs_{k-1}^L} - \frac{cvs_k^O - cvs_{k-1}^O}{cvs_{k-1}^O} \right| \times 100$$

3.1.3 Incohérence des sens d'évolution

Enfin, on mesure ici la fréquence à laquelle les deux versions de CVS donnent des évolutions de sens contraires, indépendamment du taux d'évolution. La proportion de points pour lesquels les sens d'évolution divergent est calculée ainsi :

$$CohS(Y_t) = \frac{1}{T-1} \sum_{k=2}^T I_{(cvs_k^L - cvs_{k-1}^L)(cvs_k^O - cvs_{k-1}^O) < 0} \times 100$$

Avec $I_{(cvs_k^L - cvs_{k-1}^L)(cvs_k^O - cvs_{k-1}^O) < 0} = 1$ si $(cvs_k^L - cvs_{k-1}^L)(cvs_k^O - cvs_{k-1}^O) < 0$, 0 sinon.

3.2 Comparaison empirique des méthodes

Dans les exemples ci-après on linéarise les séries avant comparaison pour se concentrer sur la stabilité du message engendré par les différentes méthodes de décomposition, en

ayant bien conscience que le pré-ajustement est aussi une importante source de divergence.

Les estimations ont été faites avec le logiciel JDemetra+ qui comprend un algorithme complet de désaisonnalisation fondé sur les modèles structurels, avec correction de jours ouvrables et outliers [12] accessible via le package R `rdj3sts`[18]

3.3 Séries mensuelles d'indices

On s'intéresse dans un premier temps aux séries d'indices à la production industrielle (Ipi), parmi celles-ci les séries fortement saisonnières avec une composante irrégulière faible sont fréquentes. On prend comme exemple l'indice sur l'exploitation de gravières et sablières, extraction d'argiles et de kaolin (RF0812), représentée sur la figure 4.

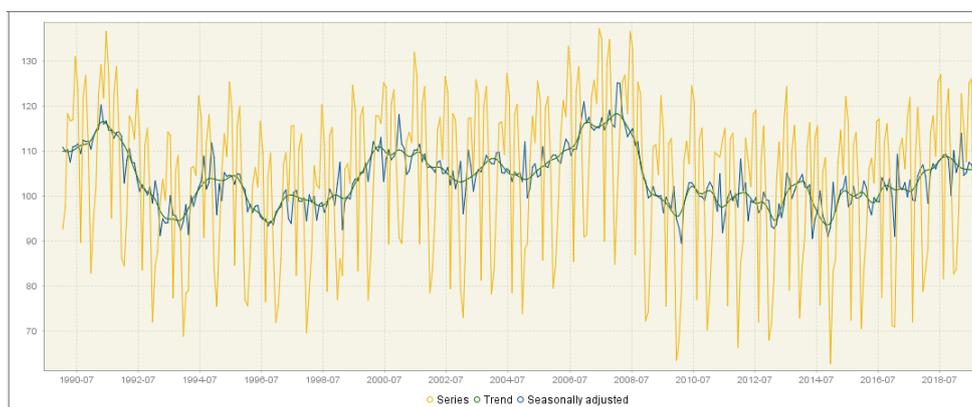


FIGURE 4 – Série Ipi RF0812 : brute (en jaune), cvs X-13 (en bleu), tendance (en vert)

Les séries cvs estimées sur longue période par Tramo-Seats ou en utilisant les BSM sont pratiquement identiques, alors que la cvs issue de X-13 présente un profil légèrement différent, comme le montre la figure 5 où ces séries sont représentées sur les cinq dernières années. La proximité entre Tramo-Seats et le BSM n'est pas surprenante car la structure d'auto-corrélation est ici proche d'un modèle Arima airline $(0,1,1)(0,1,1)$.

Pour cette série, les écarts en niveau sont inférieurs à 1% pour les trois méthodes (Figure 6). Les cvs estimées sur période longue et optimale ont des sens de variation divergents pour environ 5% des points si l'on utilise X-13 et 13% des points pour Tramo-Seats et les BSM, soit 12 points sur 95. Les résultats obtenus avec X-13 sont les moins sensibles à la longueur de la période d'estimation.

Toujours pour ce profil de séries fortement saisonnières, on peut trouver des exemples où les modèles structurels offrent une plus grande stabilité que Tramo-Seats. C'est le cas pour l'indice relatif à l'exploitation de laiteries et fabrication de fromage (RF1051) qui a une structure d'auto-corrélation plus complexe $(3,1,1)(0,1,1)$. Toutefois, les résultats obtenus avec X-13 restent dans ce cas aussi les moins sensibles à la longueur de la période d'estimation, comme le montre le tableau correspondant (figure 7).

3.4 Séries mensuelles de soldes d'opinion

On examine dans un deuxième temps des soldes d'opinion de l'enquête de conjoncture dans l'industrie, que l'on désaisonnalise également avec les trois méthodes sous revue, sur

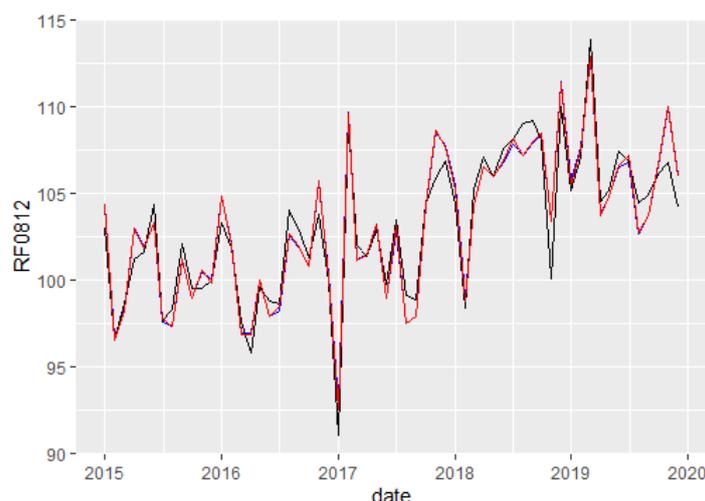


FIGURE 5 – Série Ipi RF0812 : cvs X-13 (en noir), cvs Tramo-Seats (en rouge) et BSM (en bleu)

Méthode	EN	ETC	CohS	EN 3ans	ETC 3ans	CohS 3ans
X-13	0.43	0.57	5.26	0.10	0.16	0.00
BSM	0.65	0.80	12.63	0.46	0.71	8.33
Tramo-Seats	0.78	1.11	12.63	0.60	1.01	8.33

FIGURE 6 – Série Ipi RF0812 : Indicateurs de stabilité entre estimation longue et optimale

Méthode	EN	ETC	CohS	EN 3ans	ETC 3ans	CohS 3ans
X-13	0.28	0.36	4.21	0.10	0.16	2.78
BSM	0.27	0.37	6.32	0.24	0.31	11.11
Tramo-Seats	0.29	0.37	9.47	0.26	0.28	16.67

FIGURE 7 – Série Ipi RF1051 : Indicateurs de stabilité entre estimation longue et optimale

la période longue 1990-2019 et sur une période optimale 2012-2019. Ces séries présentent des profils très différents des séries d'indices, comme l'illustre la figure 8 qui représente le solde d'opinion relatif à la production passée des trois derniers mois dans les "autres industries" (C5 en A17).

Ce solde d'opinion suit un modèle Arima (3,1,1)(0,1,1). Les estimations sur longue période présentent des profils différents selon les méthodes (figure 9) et l'on note que pour cette série c'est l'estimation avec BSM qui donne les résultats les plus instables en termes de message conjoncturel en fin de période (figure 10).

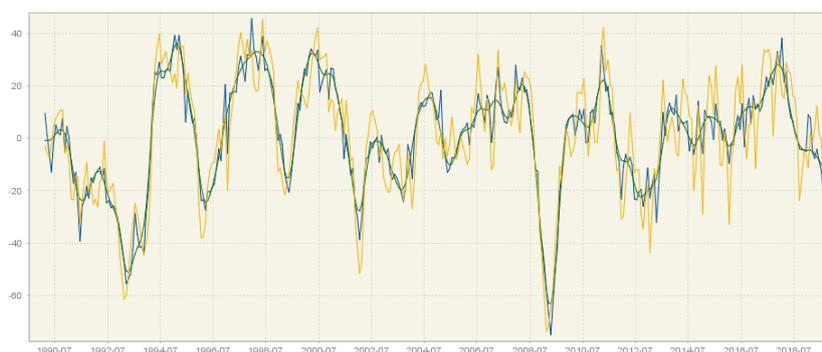


FIGURE 8 – Solde production passée secteur C5 : brute (en jaune), cvs X-13 (en bleu), tendance (en vert)

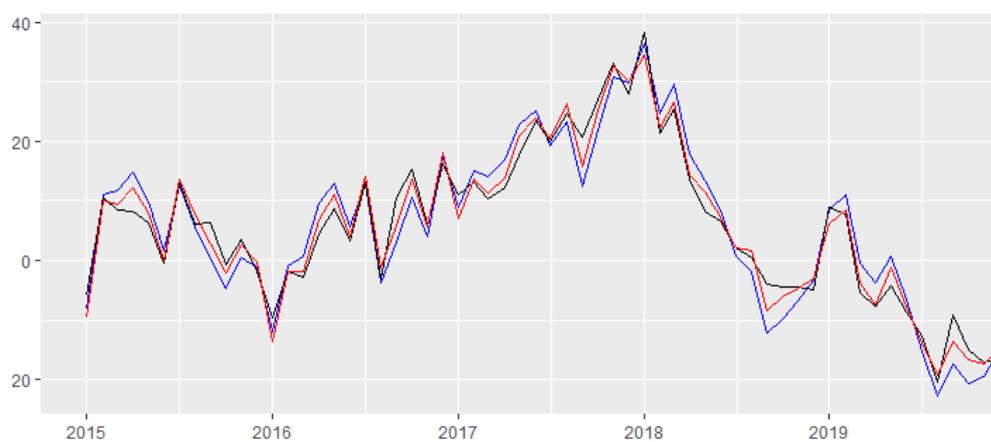


FIGURE 9 – Solde production passée secteur C5 : cvs X-13 (en noir), cvs Tramo-Seats (en rouge) et BSM (en bleu)

Le tableau des indicateurs de stabilité a été modifié par rapport aux séries d'indices car les soldes d'opinion peuvent prendre des valeurs proches de zéro ce qui rend les écarts en niveau relatifs (EN) peu informatifs, on les remplace par des écarts absolus moyens. Pour cette même raison on ne regarde pas les taux de variation mais juste la proportion de points présentant des sens de variation différents, qui est ici nettement plus élevé pour les BSM.

Méthode	EN	CohS	EN 3ans	CohS 3ans
X-13	0.81	4.21	0.33	2.78
BSM	2.86	16.84	2.55	16.67
Tramo-Seats	1.60	9.47	0.99	8.33

FIGURE 10 – Solde production passée C5 : Indicateurs de stabilité entre estimation longue et optimale

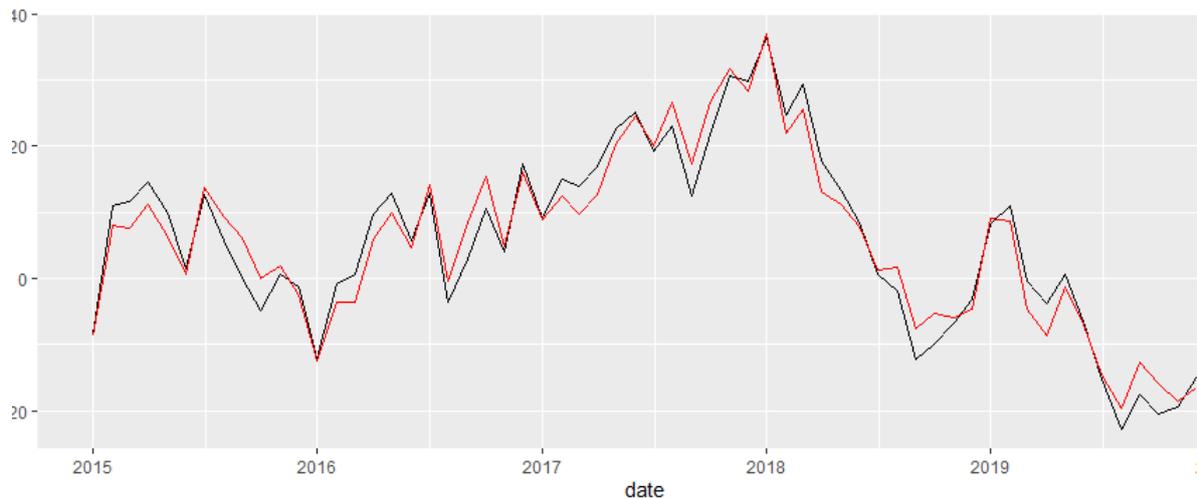


FIGURE 11 – Solde production passée C5 : cvs BSM depuis 1990 en noir et cvs BSM depuis 2012

Nos investigations empiriques confirment une plus grande stabilité de X-13 par rapport à Tramo-Seats si l'on compare les deux versions de la cvs en fin de période, en accord avec les études citées plus haut. L'apport des modèles structurels dans ce domaine n'est pas univoque : leurs estimations sont proches de celles de Tramo-Seats pour certaines structures d'autocorrélation (airline) mais peuvent aussi donner des résultats plus ou moins divergents pour d'autres. En tout cas leur sensibilité à la période d'estimation reste supérieure à celle de X-13.

Conclusion

Les modèles structurels pour les séries temporelles offrent un cadre unifié de traitement de la saisonnalité, où la dynamique des composantes inobservables est explicitement modélisée. Le pré-ajustement, correction des effets de calendrier et détection d'outliers, et la décomposition y sont traités en une seule étape. La méthode d'estimation récursive fondée sur le filtre de Kalman est particulièrement adaptée pour traiter les séries longues avec des effets variables dans le temps et éventuellement des ruptures. Un algorithme de désaisonnalisation complet reposant sur ce type de modélisation est disponible dans le logiciel JDemetra+.

Un examen empirique, illustré ici par quelques faits stylisés, montre que le message conjoncturel en fin de période est sensible à la longueur de la période d'estimation, quelque soit la méthode utilisée et ce même si l'on s'affranchit au préalable des problèmes de pré-ajustement et de la correction des effets de calendrier. Toutefois une décomposition non paramétrique (X-13), reposant sur des fenêtres mobiles d'estimation donne des résultats plus stables que les méthodes reposant sur une modélisation, structurelle ou Arima, de la série. Malgré son coût de mise en oeuvre, la coupure des séries longues semble difficilement évitable au risque de délivrer un message conjoncturel erroné.

Bibliographie

- [1] Hien PHAM et Alain QUARTIER-LA-TENTE. “Désaisonnaliser les séries très longues par sous-période”. In : *Institut National de la Statistique et des Études Économiques* (2018). URL : http://www.jms-insee.fr/2018/S05_2_ACTEv3_PHAM_JMS2018.pdf.
- [2] EUROSTAT. *ESS Guidelines on Seasonal Adjustment*. Rapp. tech. Eurostat Methodologies et Working Papers, European Commission, 2015. DOI : 10.2785/317290. URL : <http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-15-001>.
- [3] Andrew C HARVEY. *Forecasting, structural time series models and the Kalman filter*. Rapp. tech. 1989.
- [4] JDEMETRA+. *JDemetra+ software webpage*. Rapp. tech. 2020. URL : <https://github.com/jdemetra/jdemetra-app>.
- [5] Anna SMYK et Alice TCHANG. *R Tools for JDemetra+, seasonal adjustment made easier*. Rapp. tech. Documents de travail de méthodologie statistique, Insee, 2021. URL : <https://www.insee.fr/en/statistiques/5019812>.
- [6] Dominique LADIRAY et Benoit QUENNEVILLE. “Comprendre la méthode X11”. In : *Institut National de la Statistique et des Études Économiques* (1999). URL : <https://www.census.gov/ts/papers/x11doc.pd>.
- [7] Dominique LADIRAY et Alain QUARTIER-LA-TENTE. “Du bon usage des modèles Reg-Arima en désaisonnalisation”. In : *Institut National de la Statistique et des Études Économiques* (2018). URL : http://www.jms-insee.fr/2018/S05_1_ACTEv3_QUARTIERLATENTE_JMS2018.pdf.
- [8] Victor GOMEZ et Agustin MARAVALL. “Programs TRAMO (Time series regression with arima noise, missing observations, and outliers) and SEATS (Signal extraction in arima time series). Instructions for the User”. In : *Documento de Trabajo* 9628 (1996), p. 56.
- [9] Raoul DEPOUTOT et Christophe PLANAS. “Comparing seasonal adjustment and trend extraction filters with application to a model-based selection of X11 linear filters”. In : *Tech. Paper TP361, Joint Research Centre, Ispra, Italy* (1998).
- [10] Christian GOURIÉROUX et Alain MONFORT. *Séries temporelles et modèles dynamiques*. Economica, 1995.
- [11] Tommaso PROIETTI. “Comparing seasonal components for structural time series models”. In : *International Journal of Forecasting* 16.2 (2000), p. 247-260.
- [12] Jean PALATE et Anna SMYK. “Outlier detection and seasonality breaks with JDemetra+ 3.0”. In : *Conference on The New Techniques and Technologies for Statistics (NTTS)*. 2021.
- [13] Ketty ATTAL-TOUBERT et al. “Moving trading day effects with X-13 Arima-Seats and Tramo-Seats”. In : *Handbook on Seasonal Adjustment* (2018). URL : ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-001.
- [14] James DURBIN. “The foreman lecture : The state space approach to time series analysis and its potential for official statistics (with discussion)”. In : *Australian & New Zealand Journal of Statistics* 42.1 (2000), p. 1-23.

- [15] G MAZZI et G SAVIO. “Seasonal adjustment of short series”. In : *Proceedings of Statistics Canada Symposium*. 11-522. 2003, X1E.
- [16] Karsten WEBEL. “A data-driven selection of an appropriate seasonal adjustment approach”. In : (2016).
- [17] Dario BUONO, Enrico INFANTE et Gian Luigi MAZZI. “Short vs long time series : An empirical analysis Model”. In : *Handbook on Seasonal Adjustment* (2018). URL : ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-GQ-18-001.
- [18] Jean PALATE. *rjd3sts : seasonal adjustment with structural models*. URL : <https://github.com/palatej/rjd3sts>.